The EMBO Journal vol.6 no.5 pp.1429–1439, 1987

The yeast ubiquitin genes: a family of natural gene fusions

Engin Özkaynak, Daniel Finley, Mark J.Solomon[1] and Alexander Varshavsky

Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

[1]Present address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143, USA

Communicated by K.Marcker

Ubiquitin is a 76-residue protein highly conserved among eukaryotes. Conjugation of ubiquitin to intracellular proteins mediates their selective degradation *in vivo*. We describe a family of four ubiquitin-coding loci in the yeast *Saccharomyces cerevisiae*. *UBI1*, *UBI2* and *UBI3* encode hybrid proteins in which ubiquitin is fused to unrelated ('tail') amino acid sequences. The ubiquitin coding elements of *UBI1* and *UBI2* are interrupted at identical positions by non-homologous introns. *UBI1* and *UBI2* encode identical 52-residue tails, whereas *UBI3* encodes a different 76-residue tail. The tail amino acid sequences are highly conserved between yeast and mammals. Each tail contains a putative metal-binding, nucleic acid-binding domain of the form Cys-X_{2-4}-Cys-X_{2-15}-Cys-X_{2-4}-Cys, suggesting that these proteins may function by binding to DNA. The fourth gene, *UBI4*, encodes a polyubiquitin precursor protein containing five ubiquitin repeats in a head-to-tail, spacerless arrangement. All four ubiquitin genes are expressed in exponentially growing cells, while in stationary-phase cells the expression of *UBI1* and *UBI2* is repressed. The *UBI4* gene, which is strongly inducible by starvation, high temperatures and other stresses, contains in its upstream region strong homologies to the consensus 'heat shock box' nucleotide sequence. Elsewhere we show that the essential function of the *UBI4* gene is to provide ubiquitin to cells under stress.

Key words: ubiquitin/multigene family/protein turnover/TFIIIA motif/yeast

Introduction

Ubiquitin, a 76-residue protein, is apparently present in all eukaryotic cells, and is extremely conserved in evolution, with ubiquitin variants from animals, plants and yeast differing from each other in three or fewer residues out of 76 (Schlesinger and Goldstein, 1975; Gavilanes *et al.*, 1982; Dworkin-Rastl *et al.*, 1984; Özkaynak *et al.*, 1984; Bond and Schlesinger, 1985; Wiborg *et al.*, 1985; Vierstra *et al.*, 1986). Ubiquitin exists in cells either free or covalently joined via its carboxyl-terminal Gly residue to a variety of cytoplasmic, nuclear and cell surface proteins (Busch and Goldknopf, 1981; Chin *et al.*, 1982; Levinger and Varshavsky, 1982; Hershko *et al.*, 1983; Siegelman *et al.*, 1986; Yarden *et al.*, 1986). Recent biochemical and genetic evidence strongly suggests that conjugation of ubiquitin to intracellular proteins is essential for their selective degradation (reviewed by Finley and Varshavsky, 1985; Hershko and Ciechanover, 1986; see also Bachmair *et al.*, 1986; Finley *et al.*, 1987).

The exceptional evolutionary conservation of the ubiquitin amino acid sequence, and the close similarity between ubiquitin-specific enzymes from yeast and mammals (S.Jentsch, J.McGrath, and A.Varshavsky, unpublished data) suggest that insights gained from molecular genetic analyses of yeast ubiquitin should be relevant to other eukaryotes as well. We have previously reported isolation from the yeast *Saccharomyces cerevisiae* of a ubiquitin-coding gene that was found to encode a polyubiquitin precursor protein (Özkaynak *et al.*, 1984). Recently, several groups have reported the cloning of ubiquitin-coding DNA sequences from a variety of eukaryotes. Polyubiquitin genes have been identified in all five eukaryotic species examined, from yeast to mammals (Dworkin-Rastl *et al.*, 1984; Özkaynak *et al.*, 1984; Bond and Schlesinger, 1985, 1986; Wiborg *et al.*, 1985; St. John *et al.*, 1986).

In the present work we describe the family of ubiquitin-coding loci in *S. cerevisiae*. One remarkable feature of this family is that all four of the yeast ubiquitin genes encode ubiquitin fusions, either to itself (the polyubiquitin gene) or to unrelated amino acid sequences. We have also found (see Finley *et al.*, 1987), through the analysis of yeast strains with an engineered deletion of the polyubiquitin gene, that this gene is specifically required for resistance of cells to stress.

Results

Multiple ubiquitin-coding genes in yeast

We have previously cloned and partially sequenced a gene that encodes the yeast polyubiquitin precursor protein (Özkaynak *et al.*, 1984). A low-stringency Southern hybridization of electrophoretically fractionated genomic DNA with the polyubiquitin DNA probe sugests the presence of additional ubiquitin-coding loci in *S. cerevisiae* (Figure 1A, lanes a,b). The major hybridizing band in either *Hin*dIII or *Hin*dIII/*Eco*RI digests of genomic DNA corresponds to the polyubiquitin gene, whereas the minor, ~1.8 kb and ~3.5 kb bands in Figure 1A, lane b are not accounted for by this locus, suggesting that the minor bands correspond to additional ubiquitin-coding genes. Genomic DNA in the regions of ~1.8 kb and ~3.5 kb bands (Figure 1A, lane d) was purified by gel electrophoresis and cloned into an M13 vector. DNA clones containing ubiquitin-specific inserts were then isolated and sequenced (see Materials and methods). This approach yielded three new ubiquitin-coding genes, *UBI1*, *UBI2* and *UBI3* (Figures 2–4). *UBI4* designates the previously identified polyubiquitin gene (Figure 5). One striking feature of these genes is that while all of them contain ubiquitin-coding sequences, none of them encodes mature ubiquitin (Figure 6). Although the *UBI1*–*UBI4* genes encode identical amino acid sequences of ubiquitin (Figures 2–5), they differ significantly at the nucleotide sequence level (see below). Low-stringency Southern hybridization of genomic DNA with probes spanning each of the four ubiquitin-coding loci does not reveal any cross-hybridizing sequences that cannot be accounted for by the already identified ubiquitin genes (Figure 1, panels B–E, and data not shown). Thus it is likely that we have identified all of the ubiquitin-coding genes in *S. cerevisiae*.

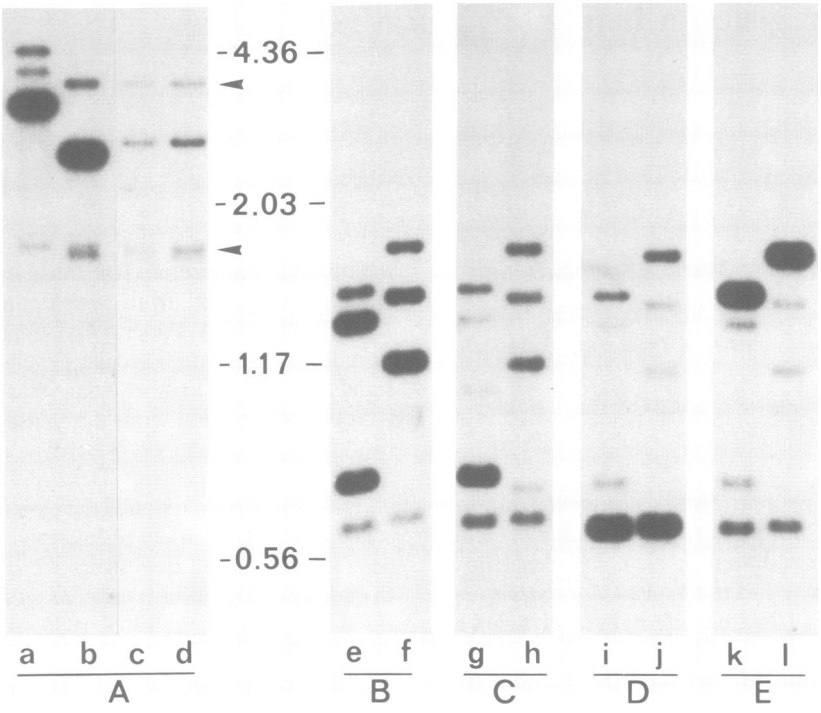


Fig. 1. Ubiquitin-coding loci in *S. cerevisiae*. Hybridizations with cloned DNA probes were carried out as described in Materials and methods. DNA in all of the lanes except c and d is from strain S288C (Sherman *et al.*, 1981). DNA in lanes c and d is from *ubi4*-Δ strains SUB16 and SUB17, respectively, in which the *UBI4* gene had been deleted (see Finley *et al.*, 1987 for a detailed description of these strains). **Lane a**, *Hin*dIII digest; **lanes b–d**, *Hin*dIII/*Eco*RI double digest; **lanes e, g, i** and **k**, *Ssp*I digest; **lanes f, h, j** and **l**, *Hae*III/*Hin*cII double digest. **Panels A** and **E**: hybridization with a ~1.3-kb *Bst*X1/*Bcl*I fragment of the *UBI4* gene (Figure 5) that contains only ubiquitin-coding repeats. **Panel B**: hybridization with a 0.46-kb *Eco*RV/*Acc*I fragment of the *UBI1* gene (Figure 2) that contains both the ubiquitin- and tail-coding sequences of *UBI1*. **Panel C**: hybridization with a 0.43-kb *Xmn*I/*Acc*I fragment of the *UBI2* gene (Figure 3) that contains both the ubiquitin- and tail-coding sequences of *UBI2*. **Panel D**: hybridization with a 0.57-kb *Ssp*I/*Hin*cII fragment of the *UBI3* gene that contains both the ubiquitin- and tail-coding sequences of *UBI3*. The indicated DNA sizes of marker DNA bands (not shown) are in kilobase pairs. Arrowheads indicate the ~1.8-kb and ~3.5-kb bands in lanes b–d that correspond to *UBI1–UBI3* loci (see Materials and methods and main text). Weak bands at ~1.02 kb in lane g and ~0.75 kb in lane h, that cannot be accounted for by the *UBI1–UBI4* genes, have been found in additional experiments to be due to an impurity in the *UBI2*-specific probe used (data not shown).

UBI1 contains an intron and encodes ubiquitin fused to an unrelated amino acid sequence

The nucleotide sequence of the *UBI1* gene is shown in Figure 2. The amino acid sequence of the ubiquitin portion of the fusion protein encoded by *UBI1* is identical to the amino acid sequences of ubiquitin encoded by *UBI2*, *UBI3* and by all five of the ubiquitin-coding repeats of *UBI4* (Figures 2–5). At the same time, there are significant differences at the nucleotide sequence level between the ubiquitin-coding elements of *UBI1–UBI4*. For instance, the ubiquitin-coding portions of *UBI1* and *UBI3* (Figures 2 and 4) differ at ~8% of the base positions, with most of the differences confined to the third positions in the corresponding codons. Moreover, the ubiquitin-coding element of *UBI1*, in contrast to those of *UBI3* and *UBI4*, is interrupted within the third codon by a 434-bp intron (Figure 2). The splice junctions of the *UBI1* intron are typical of those in yeast introns, with only one nucleotide at the 5′ junction, marked by a small dot in Figure 2, diverging from the strongly conserved consensus sequence of this site in *S. cerevisiae* (Langford and Gallwitz, 1983; Teem and Rosbash, 1983; Leer *et al.*, 1984; Green, 1986; Vijayraghavan *et al.*, 1986). The TACTAAC 'box', essential for splicing in *S. cerevisiae*, is also present within the intron (Figure 2). The *UBI1* intron has no other obvious homologies to either previously identified yeast introns or to the intron of the *UBI2* gene, which encodes an identical fusion protein (Figures 2 and 3). While most of the genes for yeast ribosomal proteins contain introns, few other yeast genes do (Leer *et al.*, 1984; Teem *et al.*, 1984; Green, 1986). Both the scarcity of yeast introns and their non-random distribution between genes of different functional classes suggest that some yeast introns serve a regulatory function. In one case this expectation has been confirmed (Dabeva *et al.*, 1986). No evidence on the possible function of introns in the *UBI1* and *UBI2* genes is currently available.

The fusion protein encoded by the *UBI1* gene is 128 residues long and consists of ubiquitin followed by an unrelated 52-residue sequence (Figure 2). Although there is no direct evidence that the fusion protein is efficiently cleaved *in vivo* between residues 76 and 77 (Figure 2) to yield mature ubiquitin, recent insights into substrate requirements of the apparently ubiquitin-specific processing protease (Bachmair *et al.*, 1986) suggest that such cleavage does occur. In particular, it has been found that a yeast protease that cleaves the engineered ubiquitin–β-galactosidase fusion proteins *in vivo* at the ubiquitin–β-galactosidase junction is generally insensitive to the nature of the first residue of β-galactosidase at the junction (Bachmair *et al.*, 1986).

The 52-residue tail of the UBI1 fusion protein is basic, containing 31% lysine and arginine residues. A highly basic stretch of seven residues at its carboxyl terminus (Figure 2) strongly resembles a sequence motif shown previously to be required for protein localization to the nucleus (reviewed by Dingwall and Laskey, 1986). These data suggest that either the intact fusion protein or its 52-residue tail may function as a nucleic acid binding protein. This possibility is strengthened by the presence within

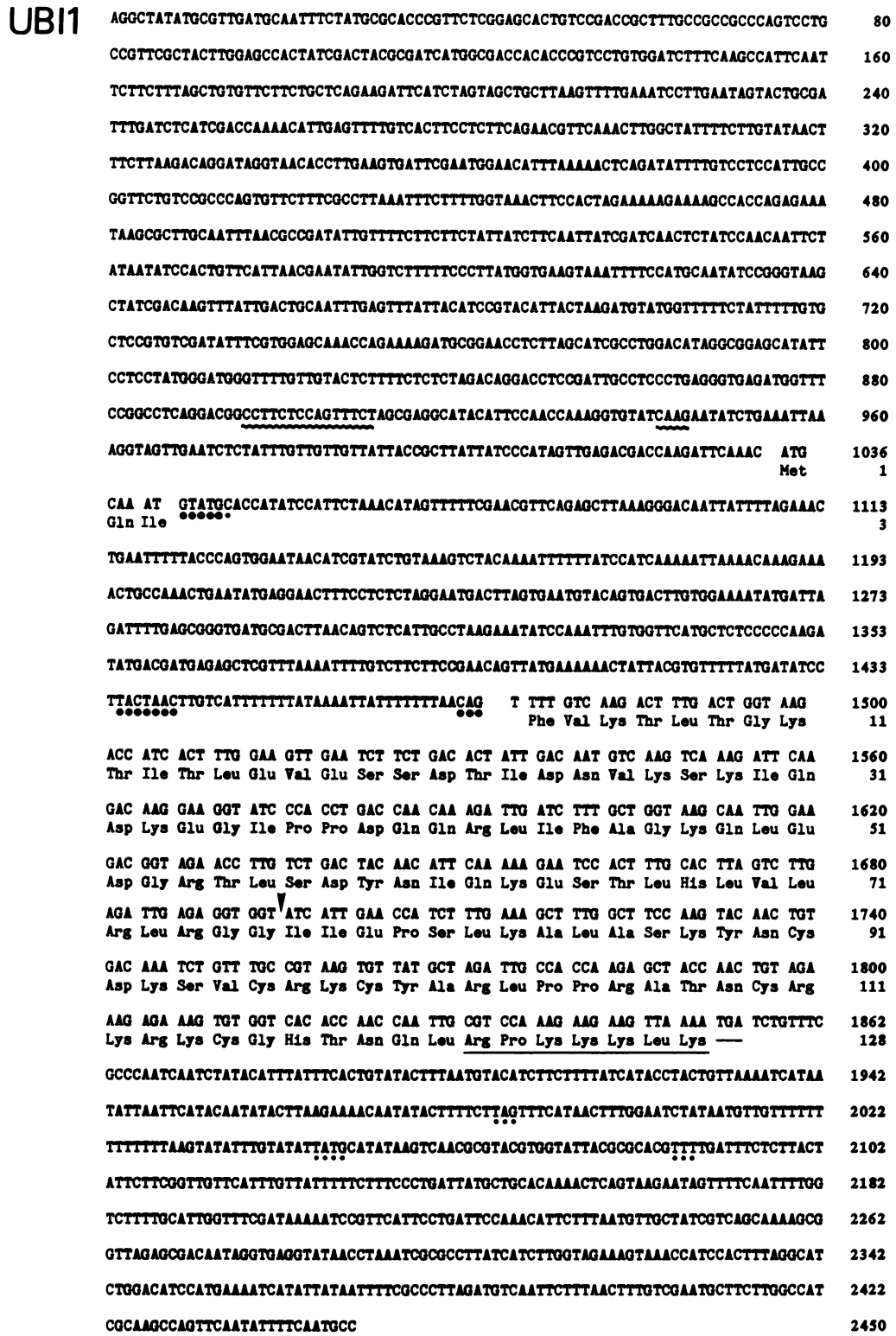


Fig. 2. Nucleotide sequence and deduced amino acid sequences of the *UBI1* locus. The TACTAAC box within the intron and the conserved intron sequences abutting the 5′ and 3′ slice junctions are marked by large dots. A nucleotide residue that does not match the consensus sequence for the 5′ splice junction (Leer *et al.*, 1984) is marked by a small dot. An arrowhead indicates the site of proteolytic cleavage that would be required to generate mature ubiquitin from the primary translation product. A stretch rich in basic amino acid residues that resembles a nuclear localization signal (Dingwall and Laskey, 1986) is underlined in the tail sequence. Wavy lines mark a bipartite nucleotide sequence motif (a CT-rich block followed, a variable distance downstream, by the sequence CAAG) common to the upstream regions of many but not all yeast genes; its function is unknown (Dobson *et al.*, 1982). A nucleotide sequence motif that resembles the consensus sequence for transcription termination in yeast (Zaret and Sherman, 1982) is marked by small dots. We have also examined the coding and flanking regions of the *UBI1*−*UBI4* genes for the presence of several other known regulatory motifs, such as sequences present upstream of the genes for ribosomal proteins (Teem *et al.*, 1984; Woudt *et al.*, 1986), sequences responsible for the cell cycle-specific expression of certain yeast genes (Nasmyth, 1985), and the known recognition sequences for regulatory proteins of the yeast mating system (Miller *et al.*, 1985; Russell *et al.*, 1986). No significant homologies to any of the above sequences have been detected in *UBI1*−*UBI4*.

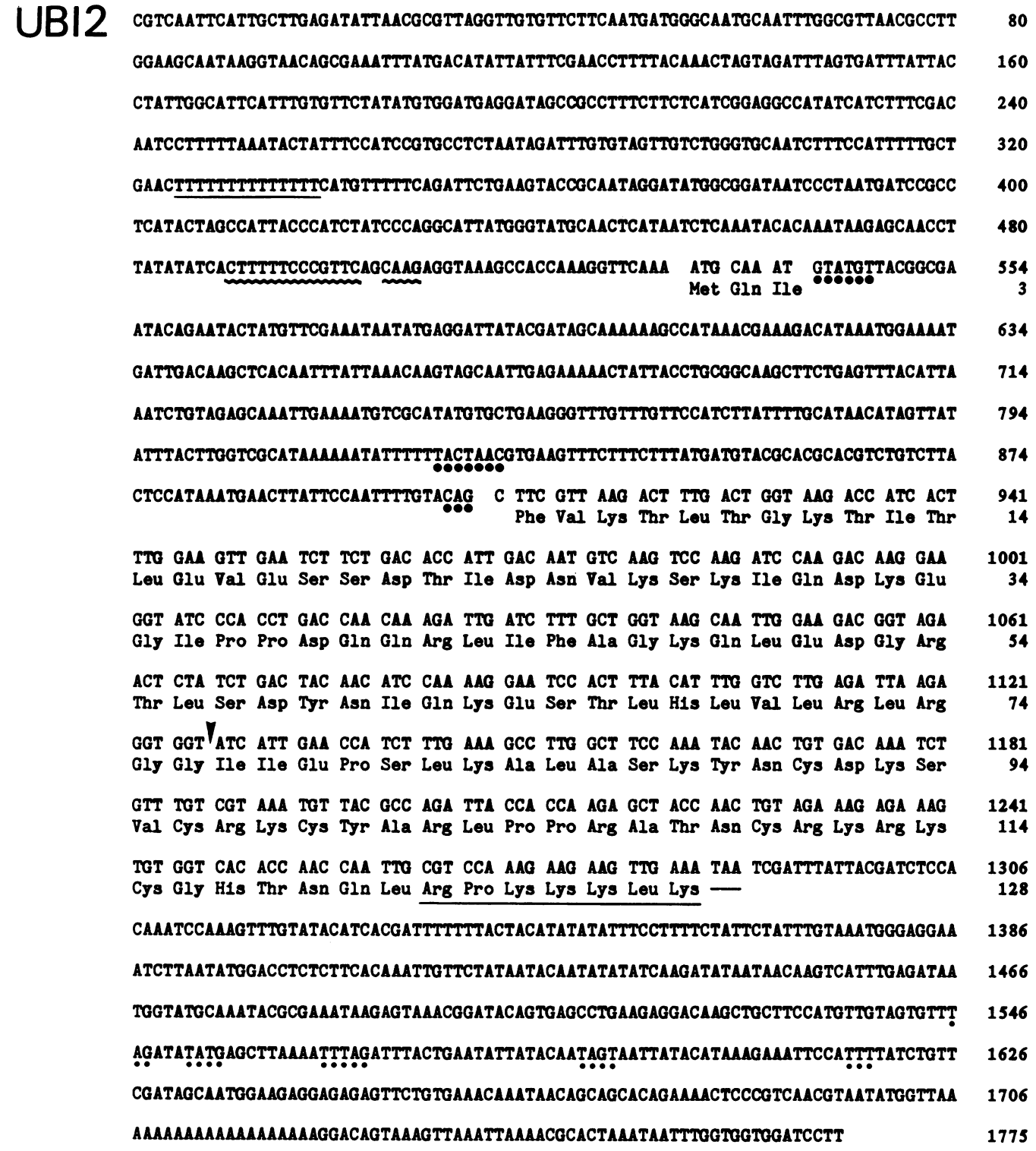


Fig. 3. Nucleotide sequence and deduced amino acid sequence of the *UBI2* locus. A stretch of 14 Ts in the upstream region (see Struhl, 1985) is underlined. Other designations are as in Figure 2.

the tail of a cysteine-containing sequence motif originally identified within a 5S RNA gene-specific transcription factor, TFIIIA (Miller *et al.*, 1985), and subsequently found in a number of other nucleic acid binding proteins (see below).

UBI2 contains an intron and encodes a ubiquitin fusion protein identical to that encoded by UBI1

The nucleotide sequence of the *UBI2* gene is shown in Figure 3. Despite the ~15% divergence at the nucleotide sequence level between the coding regions of *UBI1* and *UBI2*, they encode identical fusion proteins (Figures 2 and 3). Although the 367-bp *UBI2* intron interrupts the ubiquitin-coding sequence of *UBI2* at exactly the same position as in *UBI1*, the two introns differ in size and are not obviously homologous except for the essential sequence elements that are generally conserved among yeast in-

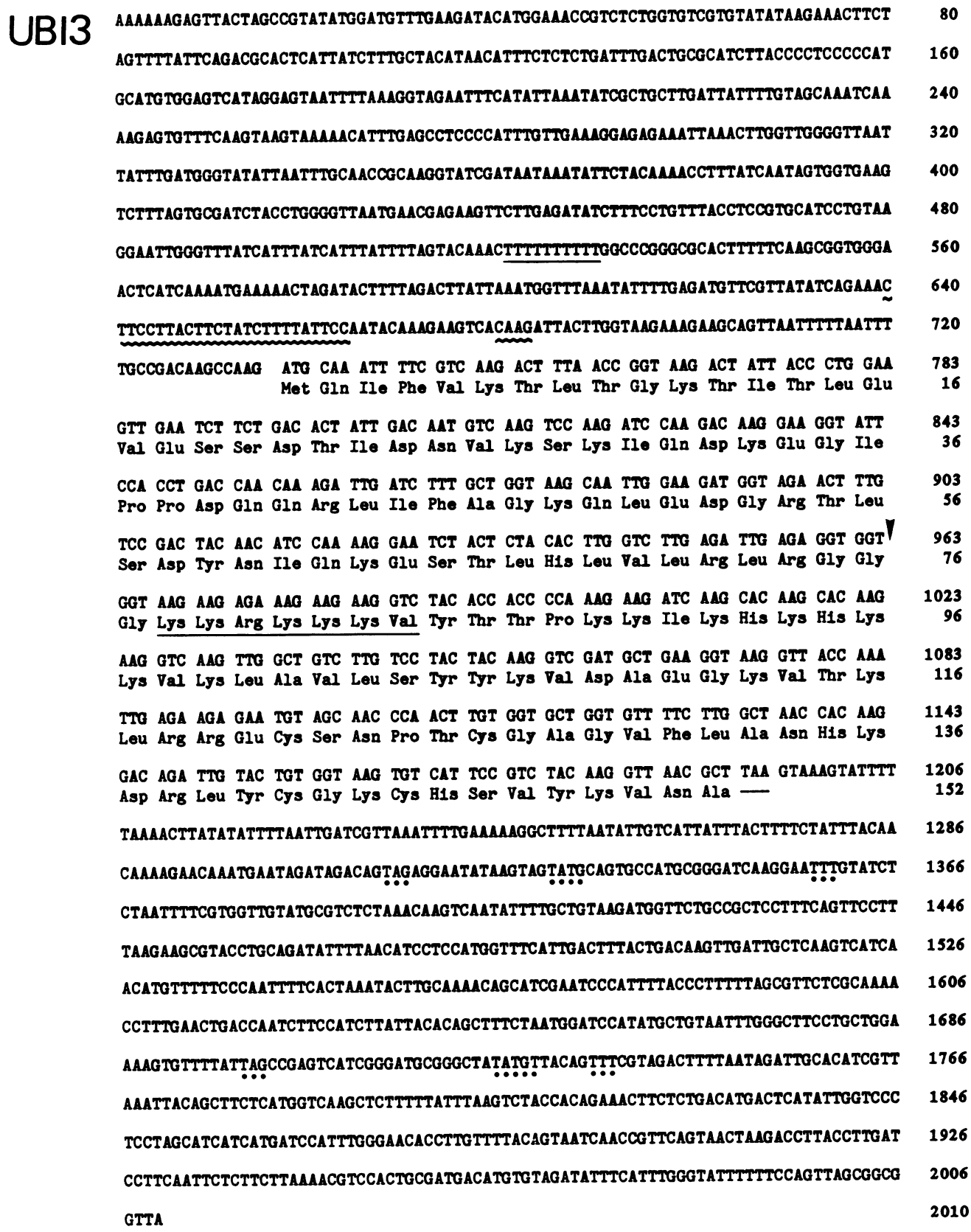


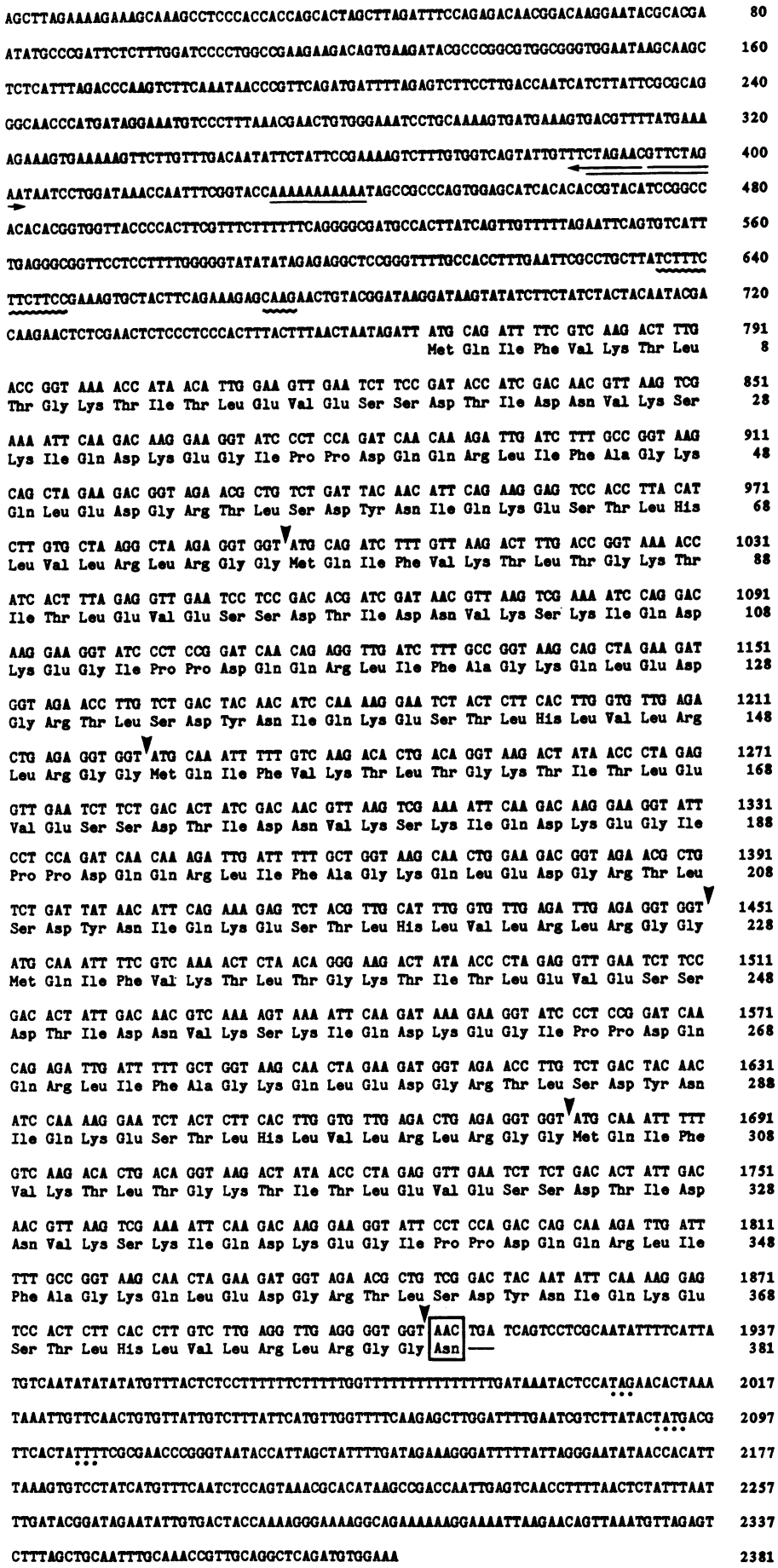
Fig. 4. Nucleotide sequences and deduced amino acid sequence of the *UBI3* locus. Designations are as in Figures 2 and 3.

trons (Figures 2, 3 and 6). The flanking regions of the *UBI1* and *UBI2* genes are also largely non-homologous, suggesting that *UBI1* and *UBI2*, while encoding identical proteins, may be differentially regulated *in vivo*.

UBI3 encodes a ubiquitin fusion protein distinct from that encoded by UBI1 and UBI2

The nucleotide sequence of the *UBI3* gene is shown in Figure 4. In contrast to the *UBI1* and *UBI2* genes, the coding region

UBI4


AGCTTAGAAAAGAAAGCAAAGCCTCCCACCACCAGCACTAGCTTAGATTTCCAGAGACAACGGACAAGGAATACGCACGA 80
ATATGCCCGATTCTCTTTGGATCCCCTGGCCGAAGAAGACAGTGAAGATACGCCCGGCGTGGCGGGTGGAATAAGCAAGC 160
TCTCATTTAGACCCAAGTCTTCAAATAACCCGTTCAGATGATTTTAGAGTCTTCCTTGACCAATCATCTTATTCGCGCAG 240
GGCAACCCATGATAGGAAATGTCCCTTTAAACGAACTGTGGGAAATCCTGCAAAAGTGATGAAAGTGACGTTTTATGAAA 320
AGAAAGTGAAAAAGTTCTTGTTTGACAATATTCTATTCCGAAAAGTCTTTGTGGTCAGTATTGTTTCTAGAACGTTCTAG 400
AATAATCCTGGATAAACCAATTTCGGTACCAAAAAAAAAAATAGCCGCCCAGTGGAGCATCACACACCGTACATCCGGCC 480
ACACACGGTGGTTACCCCACTTCGTTTCTTTTTTCAGGGGCGATGCCACTTATCAGTTGTTTTTAGAATTCAGTGTCATT 560
TGAGGGCGGTTCCTCCTTTTGGGGGTATATATAGAGAGGCTCCGGGTTTTGCCACCTTTGAATTCGCCTGCTTATCTTTC 640
TTCTTCCGAAAGTGCTACTTCAGAAAGAGCAAGAACTGTACGGATAAGGATAAGTATATCTTCTATCTACTACAATACGA 720
CAAGAACTCTCGAACTCTCCCTCCCACTTTACTTTAACTAATAGATT ATG CAG ATT TTC GTC AAG ACT TTG 791
Met Gln Ile Phe Val Lys Thr Leu 8
ACC GGT AAA ACC ATA ACA TTG GAA GTT GAA TCT TCC GAT ACC ATC GAC AAC GTT AAG TCG 851
Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Ser Ser Asp Thr Ile Asp Asn Val Lys Ser 28
AAA ATT CAA GAC AAG GAA GGT ATC CCT CCA GAT CAA CAA AGA TTG ATC TTT GCC GGT AAG 911
Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys 48
CAG CTA GAA GAC GGT AGA ACG CTG TCT GAT TAC AAC ATT CAG AAG GAG TCC ACC TTA CAT 971
Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His 68
CTT GTG CTA AGG CTA AGA GGT GGT ATG CAG ATC TTT GTT AAG ACT TTG ACC GGT AAA ACC 1031
Leu Val Leu Arg Leu Arg Gly Gly Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr 88
ATC ACT TTA GAG GTT GAA TCC TCC GAC ACG ATC GAT AAC GTT AAG TCG AAA ATC CAG GAC 1091
Ile Thr Leu Glu Val Glu Ser Ser Asp Thr Ile Asp Asn Val Lys Ser Lys Ile Gln Asp 108
AAG GAA GGT ATC CCT CCG GAT CAA CAG AGG TTG ATC TTT GCC GGT AAG CAG CTA GAA GAT 1151
Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp 128
GGT AGA ACC TTG TCT GAC TAC AAC ATC CAA AAG GAA TCT ACT CTT CAC TTG GTG TTG AGA 1211
Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg 148
CTG AGA GGT GGT ATG CAA ATT TTT GTC AAG ACA CTG ACA GGT AAG ACT ATA ACC CTA GAG 1271
Leu Arg Gly Gly Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu 168
GTT GAA TCT TCT GAC ACT ATC GAC AAC GTT AAG TCG AAA ATT CAA GAC AAG GAA GGT ATT 1331
Val Glu Ser Ser Asp Thr Ile Asp Asn Val Lys Ser Lys Ile Gln Asp Lys Glu Gly Ile 188
CCT CCA GAT CAA CAA AGA TTG ATT TTT GCT GGT AAG CAA CTG GAA GAC GGT AGA ACG CTG 1391
Pro Pro Asp Gln Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu 208
TCT GAT TAT AAC ATT CAG AAA GAG TCT ACG TTG CAT TTG GTG TTG AGA TTG AGA GGT GGT 1451
Ser Asp Tyr Asn Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly 228
ATG CAA ATT TTC GTC AAA ACT CTA ACA GGG AAG ACT ATA ACC CTA GAG GTT GAA TCT TCC 1511
Met Gln Ile Phe Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Ser Ser 248
GAC ACT ATT GAC AAC GTC AAA AGT AAA ATT CAA GAT AAA GAA GGT ATC CCT CCG GAT CAA 1571
Asp Thr Ile Asp Asn Val Lys Ser Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln 268
CAG AGA TTG ATT TTT GCT GGT AAG CAA CTA GAA GAT GGT AGA ACC TTG TCT GAC TAC AAC 1631
Gln Arg Leu Ile Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn 288
ATC CAA AAG GAA TCT ACT CTT CAC TTG GTG TTG AGA CTG AGA GGT GGT ATG CAA ATT TTT 1691
Ile Gln Lys Glu Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Met Gln Ile Phe 308
GTC AAG ACA CTG ACA GGT AAG ACT ATA ACC CTA GAG GTT GAA TCT TCT GAC ACT ATT GAC 1751
Val Lys Thr Leu Thr Gly Lys Thr Ile Thr Leu Glu Val Glu Ser Ser Asp Thr Ile Asp 328
AAC GTT AAG TCG AAA ATT CAA GAC AAG GAA GGT ATT CCT CCA GAC CAG CAA AGA TTG ATT 1811
Asn Val Lys Ser Lys Ile Gln Asp Lys Glu Gly Ile Pro Pro Asp Gln Gln Arg Leu Ile 348
TTT GCC GGT AAG CAA CTA GAA GAT GGT AGA ACG CTG TCG GAC TAC AAT ATT CAA AAG GAG 1871
Phe Ala Gly Lys Gln Leu Glu Asp Gly Arg Thr Leu Ser Asp Tyr Asn Ile Gln Lys Glu 368
TCC ACT CTT CAC CTT GTC TTG AGG TTG AGG GGT GGT AAC TGA TCAGTCCTCGCAATATTTTCATTA 1937
Ser Thr Leu His Leu Val Leu Arg Leu Arg Gly Gly Asn — 381
TGTCAATATATATATGTTTACTCTCCTTTTTTCTTTTTGGTTTTTTTTTTTTTTTGATAAATACTCCATAGAACACTAAA 2017
TAAATTGTTCAACTGTGTTATTGTCTTTATTCATGTTGGTTTTCAAGAGCTTGGATTTTGAATCGTCTTATACTATGACG 2097
TTCACTATTTTCGCGAACCCGGGTAATACCATTAGCTATTTTGATAGAAAGGGATTTTTATTAGGGAATATAACCACATT 2177
TAAAGTGTCCTATCATGTTTCAATCTCCAGTAAACGCACATAAGCCGACCAATTGAGTCAACCTTTTAACTCTATTTAAT 2257
TTGATACGGATAGAATATTGTGACTACCAAAAGGGAAAAGGCAGAAAAAAGGAAAATTAAGAACAGTTAAATGTTAGAGT 2337
CTTTAGCTGCAATTTGCAAACCGTTGCAGGCTCAGATGTGGAAA 2381

of *UBI3* lacks an intron (Figure 4). While the amino acid sequence of ubiquitin encoded by *UBI3* is identical to those encoded by *UBI1*, *UBI2* and *UBI4*, the amino acid sequence of the 76-residue UBI3 tail is quite different from the 52-residue tail of the UBI1 and UBI2 proteins (Figures 2–4). At the same time, the tails of both UBI1/UBI2 and UBI3 fusion proteins are basic, containing 31% and 29% lysine/arginine residues respectively. In addition, both tails contain putative nuclear localization sequences, although in the UBI3 protein this sequence is present at the beginning of the tail rather than at its end as in the UBI1 and UBI2 proteins (Figures 2 and 3). Moreover, a TFIIIA-like, putative nucleic acid-binding motif present in the UBI1 and UBI2 tails is also found in the UBI3 tail. As in the case of the UBI1 and UBI2 fusion proteins (see above), it is likely but not certain that the UBI3 fusion protein is cleaved *in vivo* between residues 76 and 77 to yield mature ubiquitin and a free tail protein.

UBI4 encodes a polyubiquitin fusion protein

The complete nucleotide sequence of the *UBI4* gene, determined in part previously (Özkaynak *et al.*, 1984), is shown in Figure 5. The *UBI4* gene encodes a ~43-kd primary translation product composed of five identical repeats of the ubiquitin amino acid sequence joined head-to-tail, without spacers (Figure 5). To generate mature ubiquitin from this putative precursor protein, a processing protease would have to cleave at Gly–Met junctions between the adjacent repeats (Figure 5). Although the polyubiquitin-processing protease remains to be characterized biochemically, a protease that de-ubiquitinates ubiquitin–β-galactosidase fusion proteins *in vivo* has recently been identified by Bachmair *et al.* (1986). The substrate requirements of this enzyme are those expected of a protease that processes both polyubiquitin and the other natural ubiquitin fusions encoded by the *UBI1*, *UBI2* and *UBI3* genes.

Remarkably, the last of the five ubiquitin repeats in the UBI4 protein is followed by a single Asn residue (Figures 5 and 6). Most polyubiquitin precursor proteins in higher eukaryotic species also contain an extra amino acid residue at the end of their last ubiquitin repeat, the extra residue being different in polyubiquitins from different species (reviewed by Finley and Varshavsky, 1985). By blocking the carboxyl terminus of the preceding Gly residue, the extra carboxyl-terminal residue could serve to prevent participation of unprocessed polyubiquitin in ubiquitin–protein conjugation. The properties of a recently characterized ubiquitin-specific hydrolase from mammalian cells that cleaves small adducts off the carboxyl terminus of ubiquitin (Pickart and Rose, 1986) are compatible with its involvement in removing this extra residue. Both the function of the extra residue and the identity of the protease that is resposible for its removal *in vivo* can now be addressed directly using the methods of yeast molecular genetics.

The polyubiquitin gene of *S. cerevisiae* was previously isolated from a λgt11-based library (Özkaynak *et al.*, 1984). The *UBI4* gene cloned in the present work (Figure 5) was isolated from an independently obtained, plasmid-based library. That the *UBI4* gene cloned earlier (Özkaynak *et al.*, 1984) contains six ubiquitin-coding repeats, in contrast to five such repeats in the *UBI4* gene of the present work (Figure 5), could be due either to instability of the number of ubiquitin-coding repeats upon propagation of the libraries or to a natural variation in the number of repeats from strain to strain within the same species of yeast.

Other aspects of the UBI1–UBI4 genes

A striking feature of the 5′ flanking region of *UBI4* is the presence of an 18-bp, rotationally symmetric sequence 365 bp upstream of the first codon of *UBI4* (Figure 5). The middle 14 bases of this 18-bp sequence (Figure 5) contain an exact homology to the rotationally symmetric consensus 'heat shock box' nucleotide sequence that has previously been shown to confer stress inducibility when placed upstream of heterologous promoters (Pelham, 1982; Parker and Topol, 1984; Wu, 1984; Shuey and Parker, 1985). The upstream regions of *UBI4* and *UBI3* also contain several weaker heat shock box homologies whose statistical significance is marginal. The presence of the heat shock box homologies in the 5′ flanking region of the *UBI4* gene is consistent with the stress inducibility of *UBI4* (see below), and with its specific function of providing ubiquitin to cells under stress (Finley *et al.*, 1987).

The upstream regions of *UBI2*, *UBI3* and *UBI4* (but not of *UBI1*) also contain stretches of poly(dA)·poly(dT) (underlined in Figures 3–5). Such sequences can function as upstream promoter elements for constitutive transcription in yeast (Struhl, 1985). While several putative TATA boxes (Hahn *et al.*, 1985; Nagawa and Fink, 1985; McNeil and Smith, 1986) are present upstream of the *UBI2*, *UBI3* and *UBI4* genes (Figures 3–5), there are no obvious TATA box homologies upstream of the *UBI1* gene (Figure 2).

Differential expression of UBI1–UBI4 genes

The results of Northern hybridization analysis of *UBI1–UBI4* expression in stationary-phase and exponentially growing yeast cultures, using probes specific for each of the four genes, are shown in Figure 7. The bulk of the *UBI1–UBI4* mRNAs are apparently polyadenylated, since they are retained on an oligo(dT)–cellulose column (data not shown). All four genes are expressed in exponentially growing cells, but expression of *UBI1* and *UBI2* is virtually undetectable in stationary-phase cells (Figure 7). *UBI3* is transcribed in both growing and stationary-phase cells (Figure 7, lanes e,f; note the change in size distribution of the *UBI3* mRNA species between growing and stationary-phase cells). In striking contrast, *UBI4* expression is much greater in stationary-phase cells than in growing cells (Figure 7, lanes g,h). The stress-specific expression of the *UBI4* gene accounts at least in part for the stress-specific phenotype of mutants that lack *UBI4* (Finley *et al.*, 1987).

The tails of UBI1–UBI3 proteins are conserved between yeast and higher eukaryotes

The amino acid sequence of ubiquitin is identical in mammals, frogs, fish and insects, and differs in only three of 76 residues from the sequence of yeast ubiquitin (Figure 8 and Özkaynak *et al.*, 1984). Ubiquitin from plants is even more homologous to yeast ubiquitin (two differences over 76 residues) (Vierstra *et al.*, 1986). Remarkably, the tails of the UBI1–UBI3 proteins are also strongly conserved in evolution. Specifically, the deduced amino acid sequence of the tail of the UBI3 protein is ~67% homologous to the deduced sequence of its putative human counterpart (Figure 8 and Lund *et al.*, 1985). Only a partial deduced amino acid sequence of a putative mouse counterpart

Fig. 5. Nucleotide sequence and deduced amino acid sequence of the *UBI4* locus. The underlined upstream sequence at bases 384–402 is a rotationally symmetric 18-bp sequence with an exact copy of the 14-bp consensus heat shock box sequence (Pelham, 1982; Shuey and Parker, 1985) at its center (doubly underlined). Weaker, statistically marginal homologies to the heat shock box occur also at positions 76–89, 360–367 and 721–731. The 'non-ubiquitin' Asn residue at the end of the polyubiquitin precursor protein is boxed. Other designations are as in Figures 2 and 3.

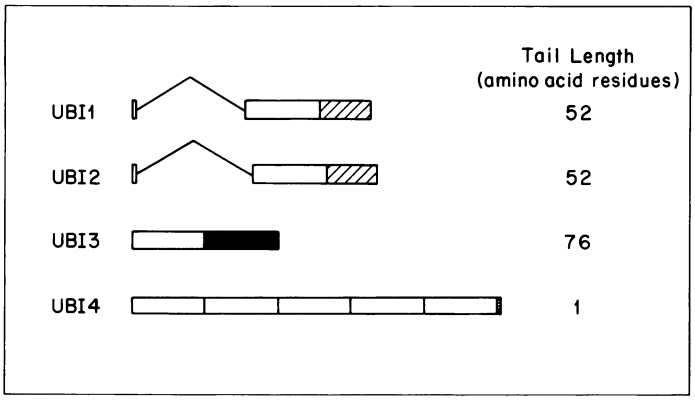


Fig. 6. Organization of yeast *UBI1*−*UBI4* genes. The relative sizes of coding elements and introns in *UBI1* and *UBI2* are drawn approximately to scale. Open blocks denote the 228-bp ubiquitin-coding elements. Two striped blocks and a dark block denote tail-coding elements in *UBI1*, *UBI2* and *UBI3* respectively. The tail amino acid sequence in the polyubiquitin product of *UBI4* consists of a single residue, Asn. See Figures 2−5 for details.

of the UBI1 and UBI2 tail sequence is known at present (Figure 8 and St. John *et al.*, 1986). In this case as well, the degree of homology in the region currently available for comparison is very high (~74%). The tail amino acid sequences are thus conserved to a remarkably high degree over great evolutionary distances, and therefore are likely to have specific if not essential functions. A search for similarities between the sequences of the UBI1−UBI3 tails and known proteins using the National Biomedical Research Foundation database and the algorithm of Lipman and Pearson (1985), has not revealed statistically significant homologies (data not shown).

The UBI1−UBI3 tails contain putative metal-binding, nucleic acid-binding domains

Recent findings by Miller *et al.* (1985) indicate that the recognition by transcription factor TFIIIA of an internal control region in the *Xenopus* 5S RNA gene is mediated by nine quasi-repeats of a ~30-residue domain that contains a coordinated Zn^{2+} ion and has several specifically arranged Cys and His residues per domain. The authors proposed that the repeating units in TFIIIA represent distinct Zn^{2+}-binding domains that interact with nucleic acids. Each of the nine domains in TFIIIA is expected to bind ~5 bp of DNA (Rhodes and Klug, 1986). A subsequent search for homologous sequences in other proteins has revealed that potential metal-binding domains occur in a number of proteins that have been implicated in nucleic acid binding (Berg, 1986; Harrison, 1986). As shown in Figure 9, the UBI1, UBI2 and UBI3 tails each contain a sequence that matches a generalized (Berg, 1986) consensus sequence of the TFIIIA motif.

Discussion

The results of this work indicate that the ubiquitin-coding sequences of yeast comprise a family of natural gene fusions. Thus, in *S. cerevisiae* ubiquitin is invariably a product of post-translational processing of precursor proteins in which ubiquitin is joined either to itself, as in the polyubiquitin (UBI4) protein, or to unrelated amino acid sequences, as in the other ubiquitin fusion proteins, UBI1−UBI3. As shown elsewhere (Finley *et al.*, 1987), at least one ubiquitin-coding locus, *UBI4* (the polyubiquitin gene), is not essential for viability of growing, unstressed vegetative cells. It is, however, absolutely required for resistance of cells to high temperatures, starvation and other stresses (Finley

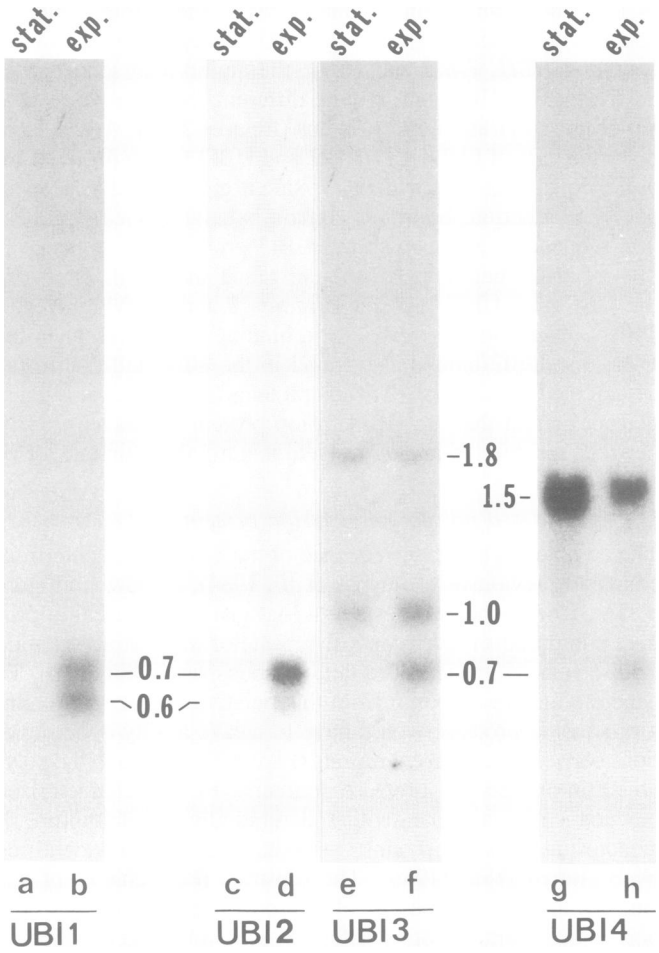


Fig. 7. Expression of *UBI1*−*UBI4* in growing and stationary cells. Total yeast RNA was electrophoresed, blotted and hybridized to DNA probes as described in Materials and methods. Equal amounts of total RNA were applied onto each lane. The same filter was used for sequential hybridizations with labeled DNA probes specific for either *UBI1* **(lanes a, b)**, *UBI2* **(lanes c, d)**, *UBI3* **(lanes e, f)** or *UBI4* **(lanes g, h)**. The *UBI1*-specific probe was an 0.34-kb *Hin*dIII/*Hin*cII fragment. The *UBI2*-specific probe was the 0.54-kb *Sty*I/*Hin*cII fragment. The *UBI3*-specific probe was an 0.51-kb *Taq*I fragment. These three probes contain 3′ flanking and tail-coding sequences of *UBI1*, *UBI2* and *UBI3*, respectively. The *UBI4*-derived probe (a *Bst*XI/*Bcl*I fragment) also cross-hybridized to ubiquitin-coding elements in the non-*UBI4* RNAs. Labels at the tops of the lanes denote RNA isolated from exponentially growing (exp.) and stationary (stat.) cultures, respectively (see Materials and methods). The indicated RNA sizes are in kilobases.

et al., 1987). The requirement for *UBI4* function in *ubi4* deletion mutants can be satisfied by an *in vitro*-constructed *UBI4* minigene that contains the flanking sequences of *UBI4* but only a single ubiquitin-coding repeat, indicating that the required function of *UBI4* is to provide mature ubiquitin monomers rather than polyubiquitin *per se* (Finley *et al.*, 1987). Since the polyubiquitin gene functions specifically during stress (Finley *et al.*, 1987), its repetitive structure may provide an additional selective advantage in reducing the metabolic cost of ubiquitin synthesis under conditions in which metabolic activities are severely compromised. In addition, formation of the spacerless polyubiquitin gene (presumably from a monoubiquitin gene) in the course of evolution may have been greatly facilitated by the prior existence of ubiquitin carboxyl-terminal proteases (Finley and Varshavsky, 1985; Bachmair *et al.*, 1986; Kanda *et al.*, 1986; Pickart and Rose, 1986). Such proteases, having evolved presumably to

```
yeast ubiquitin      : MQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQKESTLHLVLRLRGG

mammalian ubiquitin  : ------------------P----E---A----------------------------------------------

yeast UBI1,2 tail    : ...IIEPSLKALASKYNCDKSV...

mouse                : ...------RQ--Q------MR...

yeast UBI3 tail      : ...GKKRKKKVYTTPKKIKHKHKKVKLAVLSYYKVDAEGKVTKLRRECSNPTCGAGVFLANHKDRLYCGKCHSVYKVNA

human                : ...A------S------N---R--------K-----EN--ISR-----PSDE-----RM-S-F--H-----CLT-CF-KPEDK
```

Fig. 8. Conservation of the deduced amino acid sequences of the tails of UBI1−UBI3 proteins between yeast and mammals. The deduced amino acid sequences of yeast ubiquitin and of the tails encoded by *UBI1−UBI3* are taken from Figures 2−5. The deduced amino acid sequence of a human homolog of the UBI3 protein is from Lund *et al.* (1985). A portion of the mouse gene which appears to encode a counterpart of the yeast UBI1 (UBI2) protein has recently been isolated by St. John *et al.* (1986). Shown here is the currently known portion of the tail amino acid sequence encoded by the putative mouse homolog of the *UBI1*, *UBI2* genes.

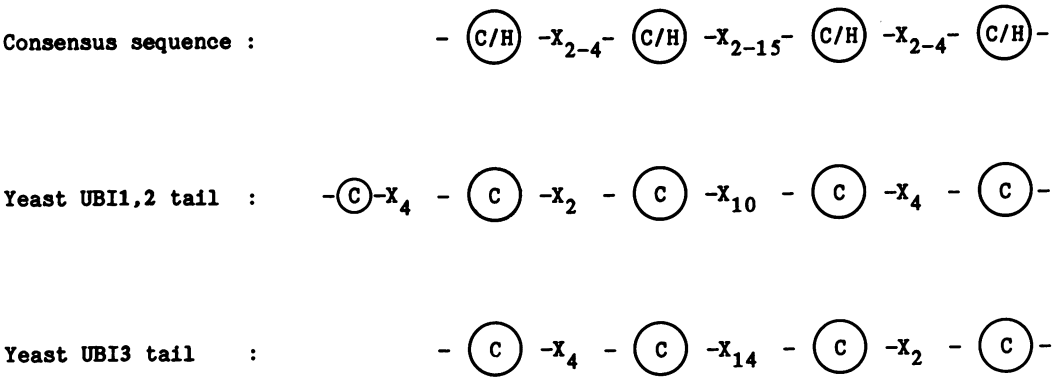


Fig. 9. Putative metal-binding, nucleic acid-binding domains in UBI1−UBI3 proteins. The consensus sequence shown (Berg, 1986) in which C and H stand for cysteine and histidine, respectively, is a generalized version of the motif originally found in the transcription factor TFIIIA of *Xenopus* and proposed to represent a Zn^{2+}-binding domain that interacts with nucleic acids (Miller *et al.*, 1985). Homologous motifs have since been detected in many other nucleic acid-binding and metal-binding proteins (reviewed by Berg, 1986; Vincent, 1986; Harrison, 1986).

recycle ubiquitin from post-translationally formed ubiquitin−protein conjugates, were thus 'pre-adapted' to process polyubiquitin and other ubiquitin fusion proteins.

Since free ubiquitin is found (at normal levels) in unstressed cells with an engineered deletion of the *UBI4* gene (Finley *et al.*, 1987), at least one of the UBI1−UBI3 proteins is expected to be processed *in vivo* to yield mature ubiquitin. These findings, and the expression patterns of the *UBI1−UBI4* genes (see Results) suggest that mature ubiquitin is derived largely from at least one of the UBI1−UBI3 fusion proteins in unstressed cells, whereas in stressed wild-type cells it is derived largely from the polyubiquitin precursor protein, UBI4. The stress-specific expression of the *UBI4* gene accounts at least in part for the stress-specific phenotype of mutants that lack *UBI4* (see Finley *et al.*, 1987 for further discussion of the function of *UBI4*).

Possible functions of the UBI1−UBI3 proteins

The genes *UBI1*, *UBI2* and *UBI3* encode hybrid proteins in which ubiquitin is fused to unrelated (tail) amino acid sequences (Figure 6). As is the case with ubiquitin itself, the tail amino acid sequences are conserved to a remarkably high degree over great evolutionary distances (Figure 8).

In addition to being basic and containing putative nuclear localization signals (see Results), the tails of UBI1, UBI2, and UBI3 proteins each contains a region that matches a generalized consensus sequence of the so-called TFIIIA motif (Figure 9). This sequence motif, which includes several specifically arranged Cys or His residues, had been originally identified in the *Xenopus* transcription factor TFIIIA, where it was found repeated nine times (Miller *et al.*, 1985). The corresponding nine domains of TFIIIA each contains a coordinated Zn^{2+} ion, and have been proposed to recognize ~5 bp of DNA per domain (Rhodes and Klug, 1986). A subsequent search for homologous sequences in other proteins has shown that the TFIIIA-like, potential metal-binding domains occur in a number of proteins, many of which have been implicated in nucleic acid binding (reviewed by Berg, 1986; Harrison, 1986; Vincent, 1986).

The tails of the UBI1−UBI3 proteins may exist *in vivo* either exclusively within the ubiquitin-containing fusion proteins (Figure 6) or in addition as free tail proteins released by cleavage at the ubiquitin−tail junction. Release of the tail proteins *in vivo* is likely, given the recently identified substrate requirements of a ubiquitin-specific processing protease (Bachmair *et al.*, 1986), and the fact that at least one of the UBI1−UBI3 proteins must be processed *in vivo* to yield mature ubiquitin (see above and also Finley *et al.*, 1987).

Whether the function of the UBI1−UBI3 tails involves specific DNA binding is not known. This possibility is being addressed by experiments in which an *in vitro*-synthesized tail protein is tested for specific DNA binding using the electrophoretic approach of Hope and Struhl (1985). Since each tail is longer than its single TFIIIA-like motif, the putative DNA-binding site of

the tail may actually encompass significantly more than ~5 bp, which is the approximate length of DNA recognized by a single TFIIIA domain (Rhodes and Klug, 1986). What might be the specific functions of the tail proteins? One possibility is that the tails function as ubiquitin-free, nucleic acid-binding proteins and act as regulators of gene expression. A free tail protein might also resemble ubiquitin in being able to form post-translational tail−protein conjugates with specific acceptor proteins. A third possibility is that each tail functions as part of the ubiquitin-containing fusion protein by binding to specific DNA sites *in vivo*, thereby greatly increasing the concentration of ubiquitin in the vicinity of such sites. Regulated proteolytic release of ubiquitin from the DNA-bound fusion protein might then result in ubiquitination events that would be limited to specific sites in the chromosomes. Finally, unprocessed ubiquitin−tail fusion proteins may function as regulators of ubiquitin-dependent protein degradation. This is suggested by the fact that, being ubiquitin−protein fusions, the UBI1−UBI3 proteins resemble post-translationally formed ubiquitin−protein conjugates, which are intermediates in ubiquitin-dependent protein degradation (for a recent discussion, see Bachmair *et al.*, 1986). In this model, the ubiquitin−tail fusion protein would be bound but not degraded by the ubiquitin-dependent protease, and as a result the degradation of post-translational ubiquitin−protein conjugates would be competitively inhibited.

The *in vitro* DNA-binding experiments described above, together with immunological studies using tail-specific antibodies, and a detailed deletion analysis of the *UBI1−UBI3* genes, currently under way, should reveal the functions of the tail proteins in yeast. The high degree of evolutionary conservation of the tail amino acid sequences suggests that the functions thus understood will be relevant in other eukaryotes as well.

Materials and methods

Isolation of DNA and RNA

Yeast culture media were prepared essentially as described by Sherman *et al.* (1981). Yeast genomic DNA was isolated from spheroplasts as described by Winston *et al.* (1983). For preparation of total RNA, yeast cultures were grown in YPD medium at 23°C and harvested either at low density (exponential cultures) or 3 h after their apparent optical density at 600 nm reached a plateau value (stationary cultures). The cells were centrifuged at 3000 *g* for 5 min, resuspended in 20 ml of 4.2 M guanidine thiocyanate at $5-10 \times 10^8$ cells/ml, immediately mixed with ~3 ml of 0.5-mm glass beads, and vortexed at the highest setting for 4 min. The lysate was centrifuged at 8000 *g* for 5 min, and RNA was purified from the supernatant by centrifugation through a cushion of 5.7 M CsCl, essentially as described by Chirgwin *et al.* (1979).

Hybridization analysis

Purified yeast DNA was digested with restriction endonucleases as recommended by their suppliers, electrophoresed in agarose gels in TAE buffer (Maniatis *et al.*, 1982), and transferred to GeneScreen filters (New England Nuclear) by the method of Reed and Mann (1985). The filters were air-dried, heated at 80°C for 1 h, and irradiated with 254 nm u.v. light (20 s at ~1.2 mW/cm^2) to cross-link DNA to the filter (Church and Gilbert, 1984). Hybridization was carried out in lucite cylinders rotating in a roller bottle apparatus at ~0.5 rev/min for ~15 h at 35°C. The hybridization solution contained 30% formamide, 4.5% SDS, 0.34 M NaCl, 1 mM Na-EDTA, 10 mg/ml bovine serum albumin, 0.16 M Na-phosphate (pH 7.0). For Northern hybridization purified RNA samples were treated with glyoxal and electrophoresed in 1% agarose gels as described by Carmichael and McMaster (1980). After electrophoresis, the gel was incubated on a rocker platform in 50 mM NaOH for 20 min, then rinsed with water and further incubated in 0.2 M Na-acetate (pH 4.0) for 30 min. RNA was then transferred by blotting in 25 mM Na-phosphate (pH 7.0) onto a GeneScreen filter. Subsequent steps were identical to those described above for Southern hybridization except that the hybridization temperature was 42°C and the hybridization solution contained 45% formamide. For additional hybridizations using the same filter, the hybridized probe was removed by incubating the filter in 0.1% SDS, 1 mM Na-EDTA, 1 mM Tris−HCl (pH 7.5) at 75°C for 1 h. The two sets of hybridization conditions listed above yielded relatively low-stringency Southern hybridization and high-stringency Northern hybridization respectively.

Cloning of S. cerevisiae ubiquitin genes

UBI1−UBI3. Yeast genomic DNA from strain SUB16 in which the polyubiquitin gene (*UBI4*) had been deleted (Finley *et al.*, 1987), was digested with both *Hin*dIII and *Eco*RI, and electrophoresed in an agarose gel. DNA was eluted from gel slices containing either ~3.5-kb DNA fragments (including the *UBI3* gene) or ~1.8-kb DNA fragments (including the *UBI1* and *UBI2* genes) (see also main text and Figure 1). A library was made from the ~3.5-kb DNA sample by ligating the DNA to *Hin*dIII/*Eco*RI-cut replicative form (RF) DNA of phage M13mp9 (Messing, 1983). Two other libraries were made by ligating the ~1.8-kb DNA fragments either to *Hin*dIII/*Eco*RI-cut M13mp9 RF DNA or to *Hin*dIII-cut and dephosphorylated M13mp9 RF DNA using standard techniques (Maniatis *et al.*, 1982). The resulting libraries were transformed into *Escherichia coli* strain JM101 (Messing, 1983), and the plaques were screened by hybridization at low stringency (Benton and Davis, 1977) with a yeast polyubiquitin (*UBI4*) gene probe (Özkaynak *et al.*, 1984; see also below) that had been labeled with ^{32}P by the method of Feinberg and Vogelstein (1984). RF DNAs were prepared from purified positive plaques of each of the three libraries. The DNA inserts were examined by gel electrophoresis and Southern hybridization with the polyubiquitin (*UBI4*) gene probe. Since the initially obtained clones of *UBI2* and *UBI3* lacked the multiple cloning site of the M13mp9 vector that is required for application of the unidirectional deletion technique (see below), the corresponding DNA inserts were electrophoretically purified, made blunt-ended (Maniatis *et al.*, 1982), and ligated to a *Hin*cII-cut M13mp9 RF DNA prior to sequencing.

Sequencing of the initially obtained *UBI1* and *UBI2* clones showed that they lacked portions of their coding and flanking sequences due to internal *Hin*dIII cuts in genomic DNA used for cloning. To clone the complete *UBI1* and *UBI2* genes, probes were prepared from unique, non-ubiquitin portions of the initial *UBI1* and *UBI2* clones. One probe (a ~1.4-kb *Hin*dIII/*Sst*I fragment) spanned the 5′ flanking region and intron of the *UBI1* gene, and the other (a ~1.1-kb *Acc*I/*Eco* RI fragment) spanned the 3′ flanking region of the *UBI2* gene. A plasmid Ycp50-based library of ~10-kb yeast genomic DNA inserts from strain S288C, carried in *E. coli* strain HB101, was obtained from Dr P.Novick (Yale University), screened with the above probes, and positive clones were purified. A *UBI1*-containing Ycp50 plasmid was cut within the *UBI1* intron with *Sst*I and at a 3′ flanking site with *Bgl*II. The resulting ~3-kb fragment was blunt-ended and ligated to *Hin*cII-cut M13mp9 RF DNA for sequencing. A *UBI2*-containing Ycp50 plasmid was cut at 5′ and 3′ flanking sites of *UBI2* with *Eco*RI and *Acc*I, respectively, and the resulting ~1.2-kb fragment was subcloned as above into M13mp9 RF DNA for sequencing.

UBI4. A yeast polyubiquitin (*UBI4*) gene probe derived from the originally obtained polyubiquitin clone carried in λgt11 (Özkaynak *et al.*, 1984) was used to screen the Ycp50-based genomic library described above. Plasmids from strongly positive colonies were purified and analysed by Southern hybridization. One plasmid thus obtained, pUB6, was digested with *Hin*dIII and the ~3-kb, *UBI4*-containing fragment was isolated for sequencing.

DNA sequencing

The unidirectional deletion procedure used to prepare the *UBI1*, *UBI2* and *UBI3* clones for DNA sequencing by the chain termination method (Sanger *et al.*, 1977) was essentially as described by Putney *et al.* (1981) and Yanisch-Perron *et al.* (1985). For sequencing of the *UBI4* gene, nine overlapping DNA fragments were generated from the ~3-kb, *UBI4*-containing DNA fragment (see above) using appropriate restriction endonucleases and the previously available partial sequence of the *UBI4* gene (Özkaynak *et al.*, 1984). The fragments were purified and subcloned into M13mp9 RF DNA for sequencing. Coding regions of *UBI1−UBI3* were sequenced on both strands. In addition, the sequence was determined at least twice in regions of initial ambiguity.

Acknowledgements

We thank Peter Novick for the Ycp50-based DNA library, Gordon Carmichael for advice on the electrophoresis of RNA, and Barbara Doran for secretarial assistance. E.Ö. is especially grateful to Ulya Özkaynak for her encouragement and help during this work. Supported by a grant to A.V. from the National Institute of General Medical Sciences (GM31530). E.Ö. was supported by a fellowship from the Charles A.King Trust and the Medical Foundation.

References

Bachmair,A., Finley,D. and Varshavsky,A. (1986) *Science*, **234**, 179−186.
Benton and Davis (1977) *Science*, **196**, 180−182.
Berg,J.M. (1986) *Science*, **232**, 485−487.
Bond,U. and Schlesinger,M.J. (1985) *Mol. Cell. Biol.*, **5**, 949−956.
Bond,U. and Schlesinger,M.J. (1986) *Mol. Cell. Biol.*, **6**, 4602−4610.

Busch,H. and Goldknopf,I. (1981) *Mol. Cell. Biochem.*, **40**, 173–187.
Carmichael,G.G. and McMaster,G.K. (1980) *Methods Enzymol.*, **65**, 380–391.
Chin,D.T., Kuehl,L. and Rechsteiner,M. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 5857–5861.
Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.F. (1979) *Biochemistry*, **18**, 5294–5299.
Church,G. and Gilbert,W. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 1991–1995.
Dabeva,M., Post-Beittenmiller,M.A. and Warner,J.R. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 5854–5857.
Dingwall,C. and Laskey,R.A. (1986) *Annu. Rev. Cell Biol.*, **2**, 366–390.
Dworkin-Rastl,E., Shrutkowski,A. and Dworkin,M.B. (1984) *Cell*, **39**, 321–325.
Dobson,M.J., Tuite,M.F., Roberts,N.A., Kingsman,A.J. and Kingsman,S.M. (1982) *Nucleic Acids Res.*, **10**, 2625–2637.
Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266–267.
Finley,D. and Varshavsky,A. (1985) *Trends Biochem. Sci.*, **10**, 343–346.
Finley,D., Özkaynak,E. and Varshavsky,A. (1987) *Cell*, **48**, 1035–1046.
Gavilanes,J.G., Gonzalez de Buitrago,G., Perez-Castells,R. and Rodriguez,R. (1982) *J. Biol. Chem.*, **257**, 10267–10270.
Green,M.R. (1986) *Annu. Rev. Genet.*, **20**, 671–698.
Hahn,S., Hoar,E.T. and Guarente,L. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8562–8566.
Harrison,S. (1986) *Nature*, **322**, 597–598.
Hershko,A. and Ciechanover,A. (1986) *Progr. Nucleic Acids Res. Mol. Biol.*, in press.
Hershko,A., Heller,H., Elias,S. and Ciechanover,A. (1983) *J. Biol. Chem.*, **256**, 8206–8214.
Hope,I.A. and Struhl,K.(1985) *Cell*, **43**, 177–188.
Kanda,F., Sykes,D.E., Yasuda,H., Sandberg,A. and Matsui,S. (1986) *Biochim. Biophys. Acta*, **870**, 64–75.
Langford,C.J. and Gallwitz,D. (1983) *Cell*, **33**, 519–527.
Leer,R.J., van-Raamsdonk-Duink,M.M.C., Hagendorn,M.J.M., Mager,W.H. and Planta,R.J. (1984) *Nucleic Acids Res.*, **12**, 6685–6700.
Levinger,L. and Varshavsky,A. (1982) *Cell*, **28**, 375–385.
Lipman,D.J. and Pearson,W.R. (1985) *Science*, **227**, 1435–1441.
Lund,P.K., Moats-Staats,B.M., Simmons,J.G., Hoyt,E., D'Ercole,A.J., Martin,F. and Van Wyk,J.J. (1985) *J. Biol. Chem.*, **260**, 7609–7613.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning. A Laboratory Manual.* Cold Spring Harbor Laboratory, New York.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.
Miller,A.M., MacKay,V.L. and Nasmyth,K. (1985) *Nature*, **314**, 598–603.
Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
McNeil,J.B. and Smith,M. (1986) *J. Mol. Biol.*, **187**, 363–378.
Nagawa,F. and Fink,G. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8557–8561.
Nasmyth,K. (1985) *Cell*, **42**, 225–235.
Özkaynak,E., Finley,D. and Varshavsky,A. (1984) *Nature*, **312**, 663–666.
Parker,C.S. and Topol,J. (1984) *Cell*, **37**, 273–283.
Pelham,H.R.B. (1982) *Cell*, **30**, 517–528.
Pickart,C.M. and Rose,I.A. (1986) *J. Biol. Chem.*, **261**, 10210–10217.
Putney,S.C., Benkovic,S.J. and Schimmel,P.R. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 7350–7354.
Reed,K.C. and Mann,D.A. (1985) *Nucleic Acids Res.*, **13**, 7207–7221.
Rhodes,D. and Klug,A. (1986) *Cell*, **46**, 123–132.
Russell,D.W., Jensen,R., Zoller,M.J., Burke,J., Errede,B., Smith,M. and Herskowitz,I. (1986) *Mol. Cell. Biol.*, **6**, 4281–4294.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schlesinger,D.H. and Goldstein,G. (1975) *Nature*, **255**, 423–424.
Sherman,F., Fink,G.R. and Hicks,J.B. (1981) *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory, New York.
Shuey,D.J. and Parker,C.S. (1985) *J. Biol. Chem.*, **261**, 7934–7940.
Siegelman,M., Bond,M.W., Gallatin,W.M., St. John,T., Smith,H.T., Fried,V.A. and Weissman,I.L. (1986) *Science*, **231**, 823–829.
St. John,T., Gallatin,W.M., Siegelman,M., Smith,H.T., Fried,V.A. and Weissman,I.L. (1986) *Science*, **231**, 845–850.
Struhl,K. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 8419–8423.
Teem,J.L. and Rosbash,M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4403–4407.
Teem,J.L., Abovich,N., Kaufer,N.F., Schwindinger,W.F., Warner,J.R., Levy,A., Woolford,J., Leer,R.J., van Raamsdonk-Duin,M.C.C., Mager,W.H., Planta,R.J., Schultz,L., Friesen,J.D., Fried,H. and Rosbash,M. (1984) *Nucleic Acids Res.*, **12**, 8295–8312.
Vierstra,R.D., Langan,S.M. and Schaller,G.E. (1986) *Biochemistry*, **25**, 3105–3108.
Vijayraghavan,U., Parker,R., Tamm,J., Iimura,Y., Rossi,J., Abelson,J. and Guthrie,C. (1986) *EMBO J.*, **5**, 1683–1695.
Vincent,A. (1986) *Nucleic Acids Res.*, **14**, 4385–4391.
Wiborg,O., Pedersen,M.S., Wind,A., Berglund,L.E., Marcker,K.A. and Vuust,J. (1985) *EMBO J.*, **4**, 755–759.
Winston,F., Chumley,F. and Fink,G.R. (1983) *Methods Enzymol.*, **101**, 211–228.
Wu,C. (1984) *Nature*, **309**, 229–234.
Woudt,L.P., Smit,A.B., Mager,W.H. and Planta,R.J. (1986) *EMBO J.*, **5**, 1037–1040.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.
Yarden,T., Escobedo,J.A., Kuang,W.-J., Yang-Feng,T.L., Daniel,T.O., Tremble,P.M., Chen,E.Y., Ando,M.E., Harkins,R.N., Francke,U., Fried,V.A., Ullrich,A. and Williams,L.T. (1986) *Nature*, **323**, 226–232.
Zaret,K.S. and Sherman,F. (1982) *Cell*, **28**, 563–573.

Received on January 26, 1987

Note added in proof

A cDNA clone from the slime mold *Dictyostelium discoideum* that encodes a ubiquitin-tail protein strongly homologous to the UBI1/UBI2 protein of yeast has recently been described by M. Westphal, A. Müller-Taubenberger and G. Gerisch (1986) *FEBS Lett.*, **209**, 92–95.

The sequence data of the present work have been submitted to the EMBL/GenBank Data Libraries under the accession number Y00061.